From Soft Decision Channel Decoding to Soft Decision Speech Decoding

Peter Vary and Tim Fingscheidt

Institut für Nachrichtengeräte und Datenverarbeitung
RWTH Aachen, Templergraben 55, D - 52056 Aachen
Tel.: +49 241 806956, Fax: +49 241 8888186,
E-Mail: vary@ind.rwth-aachen.de

Abstract

For speech transmission in digital mobile radio systems the net bit rate has to be reduced with respect to frequency economy by source encoding techniques. In a second step redundant bit rate is added by channel encoding for the purpose of error protection. As the gross bit rate is fixed, a compromise has to be taken between the speech qualities in clear and poor channel conditions. If the quality of the channel is very poor error concealment techniques have to be applied additionally at the receiver to reduce the adverse subjective effects of residual bit errors after channel decoding.

This contribution presents a universal approach to error concealment by exploiting soft decision information from the channel decoder and a priori information about speech codec parameters. The problem of joint optimization of speech coding, channel coding and error concealment is addressed. It is shown that error concealment can take over to a certain extent the task of error protection. Thus less bit rate would be required for channel coding.

1 Introduction

The general and simplified block diagram of a digital speech communication system is given in Fig. 1. Samples s of the speech signals are converted by speech encoding into a stream of bits which can be divided into groups or vectors $\underline{x}$ of a certain length. Subsequently channel encoding is applied for error protection by adding explicit redundancy and increasing the bit rate correspondingly. The sequential stream of the output bits of the channel encoder is subdivided into vectors $\underline{Y}$. The transmission part is contained in the block denoted by modulator, channel and demodulator (possibly including equalization). At the output of the demodulator two kinds of information are available: the channel encoded vectors $\hat{\underline{Y}}$ (which are in case of error free transmission identical with the vectors $\underline{Y}$) and so called channel state information CSI indicating the present quality of the channel. Modern channel decoding algorithms such as the soft output Viterbi algorithm (SOVA) [15] can exploit the channel state information to improve the performance and to provide decoder reliability information DRI about the decoded information bits, denoted in Fig. 1 by the vectors $\hat{\underline{x}}$.

In this contribution it is shown how the channel decoder reliability information (DRI) can be exploited within the process of speech decoding to reduce the subjectively annoying effects of residual bit errors. As error concealment can take over to a certain extent the task of error protection in terms of subjective speech quality, the question of joint optimization of speech coding, channel coding and error concealment will be discussed.

Due to the fact that most standards do not specify the error concealment part explicitly, there is room for new solutions to improve the decoding process even without a redesign of speech and channel coding. It turns out that besides the channel state information CSI and the decoder reliability information DRI a priori knowledge about speech codec parameters plays a major role to improve channel decoding (e.g. [14, 16, 17]) as well as speech decoding [9, 10, 11].

2 Soft Bits

The combined information of any bit and its associated quality or reliability information is called a soft bit (e.g. [14, 9, 12]). If we consider in Fig. 1 the transmission of any bit pattern or vector

$$\underline{x} = (x(0), x(1), ..., x(M-1)) \quad \text{with} \quad x(m) \in \{-1, +1\}$$

of length M we can define the equivalent channel in Fig. 2 with input $x(m)$, $m = 0, 1, ..., M-1$ and soft output

$$\{\hat{x}(m), p(m)\} \quad \text{with} \quad \hat{x}(m) \in \{-1, +1\}$$

consisting of the hard bits $\hat{x}(m)$ and the decoder reliability information which can be expressed e.g. by the instantaneous bit error rate $p(m)$. Equivalently we can define a real-valued soft bit

$$\tilde{x}(m) = \hat{x}(m) \cdot [1-2p(m)] \quad \text{with} \quad -1 \leq \tilde{x}(m) \leq +1 \, .$$

Under the assumption of equally probable source bits $\mathrm{P}(x(m) = +1) = \mathrm{P}(x(m) = -1) = 1/2$ and a sym-

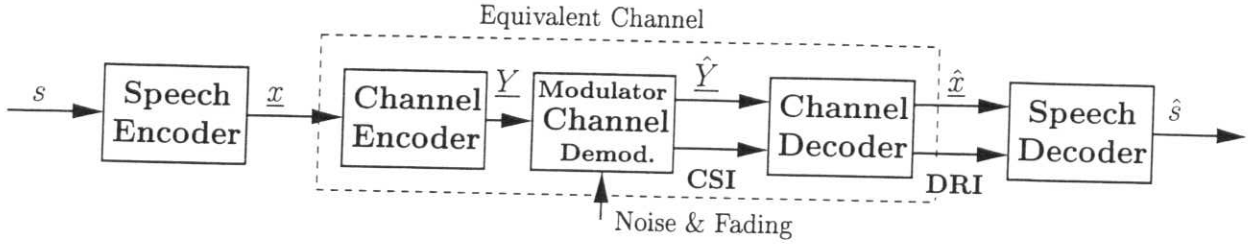

Figure 1: Block diagram of a digital speech communication system

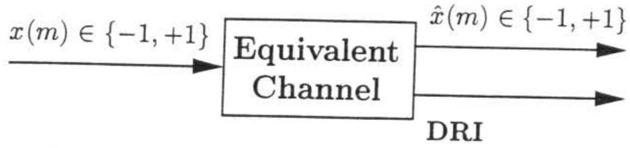

Figure 2: The equivalent channel on bit level

metric channel the soft bit equals the conditional expected value $E\{x(m) \mid \hat{x}(m)\} = \tilde{x}(m)$. In case of error free transmission the soft bit is the same as the hard bit. If the bit error rate is $p(m) = 0.5$ (i.e. no transinformation), then the soft bit becomes $\tilde{x}(m) = 0$. This is a desireable feature which can be used to achieve graceful degradation and automatic muting with increasing bit error rate. An alternative representation which is related to the metric calculation of the channel decoder is the so called logarithmic likelihood value (L-value, e.g. [14]):

$$L(m) = \hat{x}(m) \cdot \ln\left[\frac{1-p(m)}{p(m)}\right] = \hat{x}(m) \cdot \ln\left[\frac{1+|\tilde{x}(m)|}{1-|\tilde{x}(m)|}\right]$$

with

$$-\infty \leq L(m) \leq \infty .$$

The hard bit is identical with the sign of the soft bit and the sign of the L-value

$$\hat{x}(m) = \mathrm{sign}[\tilde{x}(m)] = \mathrm{sign}[L(m)] .$$

3 Channel Decoding with Soft Decision Output

An essential requirement for error concealment techniques is the reliability information DRI (see Fig. 1) provided by the channel decoder (or by the demodulator/equalizer in transmission systems without channel coding).

In the sequel it is assumed that convolutional codes with rate r are used. These codes can be decoded by taking channel state information as well as a priori knowledge on bit level or parameter level into consideration. In [14] a modified Viterbi algorithm (VA) had been proposed which uses a priori or a posteriori information about source bit probabilities. This algorithm is an extension of the soft output Viterbi algorithm (SOVA) as described in [15] and is called

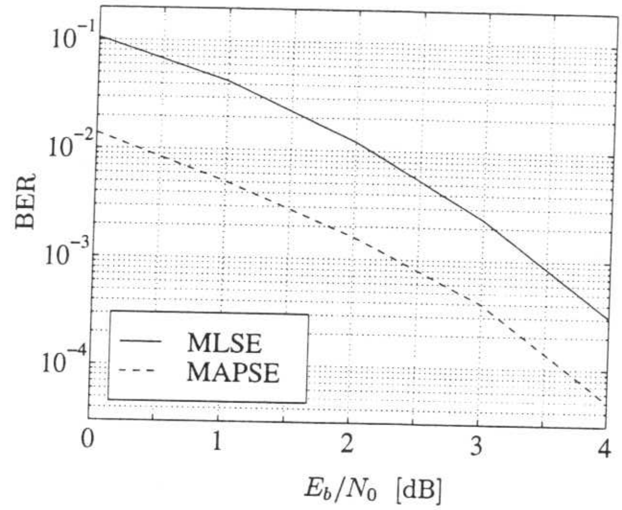

Figure 4: Residual bit error rate (BER) for maximum likelihood sequence estimation without a priori knowledge (MLSE) and maximum a posteriori sequence estimation with a priori knowledge (MAPSE) [16]; Gaussian signal, 4 bit per sample

APRI-SOVA. It performs "source controlled channel decoding". The basic objective is to minimize the bit error rate by *maximum a posteriori sequence* estimation (MAPSE). An alternative solution which is under certain circumstances more powerful but more complex had been proposed in [6]. It is based on the *maximum a posteriori symbol* criterion. This algorithm is able to take a priori knowledge on bit level into consideration. Channel decoding mechanisms exploiting a priori knowledge on parameter level are e.g. [16, 17]. Using these techniques additional gains in terms of the E_b/N_0 ratio can be achieved of about 1...2 dB.

An example is given in Fig. 4 [16]. The source encoder output is modelled by white Gaussian noise $\sigma_v^2 = 1$. This signal is uniformly quantized with 16 levels using a non-symmetric midtreat quantizer. The entropy of this source is about 2 bit per sample, the bits of the 4 bit parameter are statistically dependent from each other. The a priori knowledge is given by the non-uniform distribution of the quantized parameter.

The channel encoder of rate $r = 1/2$ has the generator polynomial coefficients $G_1 = 15$ and $G_2 = 17$ (octal). Simulation results are presented in Fig. 4 for an AWGN channel with coherent BPSK demodulation using the following algorithms:

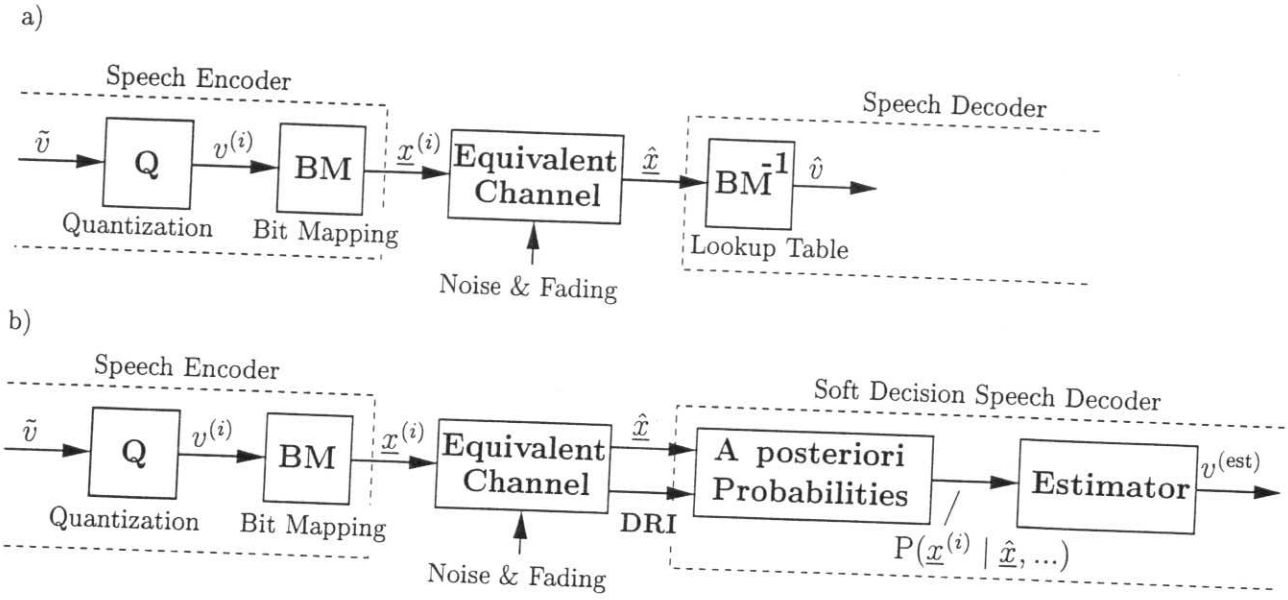

Figure 3: Quantization, bit mapping and transmission of a speech codec parameter $\tilde{v}$, decoding by ...
a) ... conventional hard decision speech decoding via table lookup (HD)
b) ... soft decision speech decoding (SD)

- MLSE: Maximum likelihood sequence estimation without a priori knowledge
- MAPSE: Maximum a posteriori sequence estimation using a priori knowledge on parameter level.

At a bit error rate of $\mathrm{BER} = 10^{-3}$ the E_b/N_0 improvement is in this example about 1 dB. The algorithm can be implemented without significant increase of complexity in comparison to the plain MLSE approach. For further details see [16]. The achievable improvement is depending on the amount of redundancy of the quantized parameter.

4 Speech Decoding with Soft Decision Input

Suppose a quantized codec parameter v is constructed according to

$$v = \sum_{m=0}^{M-1} A_m \cdot x(m) \tag{1}$$

with $x(m)$ being its representing bits and A_m are the amplitude contributions of the bits. The soft decision source decoded parameter $v^{(\mathrm{est})}$ can then easily be reconstructed or estimated by

$$\begin{aligned} v^{(\mathrm{est})} = E\{v \mid \hat{\underline{x}}\} &= \sum_{m=0}^{M-1} A_m \cdot E\{x(m) \mid \hat{x}(m)\} \qquad (2) \\ &= \sum_{m=0}^{M-1} A_m \cdot \tilde{x}(m)\ . \end{aligned}$$

Unfortunately, most quantizers do not allow a bit mapping according to eq. (1). Furthermore, it is difficult to include a priori knowledge about the statistics of the parameter in eq. (2). Finally, the estimation of the codec parameter by its expected value corresponds to a squared error criterion. For some parameters even other estimation criteria are subjectively superior. This leads to the following more general conception of soft decision source decoding.

4.1 Basic Structure

In general terms the quality of the decoded speech especially under poor channel conditions depends on the proper estimation of codec parameters. For this reason, we focus on the estimation of codec parameters rather than on the detection of individual bits. This approach is called error concealment by soft decision or soft bit speech decoding (e.g. [9]).

Let us consider the transmission of a specific codec parameter $\tilde{v} \in \mathbb{R}$ as depicted in Figure 3a. The quantized parameter $\mathbf{Q}[\tilde{v}] = v$ with $v \in \mathrm{QT}$ (QT: quantization table) is represented by the combination $\underline{x} = (x(0), x(1), ..., x(m), ..., x(M-1))$ consisting of M bits. The bits $\underline{x}(m)$ are assumed to be bipolar, i.e. $x(m) \in \{-1, +1\}$. Any bit combination $\underline{x}$ is assigned to a quantization table index i, such that we can denote the quantized parameter by $\underline{x} = \underline{x}^{(i)}$ or equivalently by $v = v^{(i)}$ with $i \in \{0, 1, ..., 2^M - 1\}$. Furthermore, we distinguish the quantities $\hat{v}, \hat{\underline{x}}$ at the receiver from those at the transmitter.

In a conventional decoding scheme as indicated in Fig. 3a the received bit combination $\hat{\underline{x}}$ is applied to "inverse bit mapping" or "inverse quantization", i.e. the parameter $\hat{v}$ is read out of a quantization table.

The proposed error concealment technique consists of the modified decoding procedure as shown in Fig. 3b: In the first step we calculate a posteriori probabilities

$$\mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{X}}_{-1}) = C \cdot \mathrm{P}(\hat{\underline{x}}_0 \mid \underline{x}_0{}^{(i)}) \cdot \sum_{j=0}^{2^M-1} \mathrm{P}(\underline{x}_0{}^{(i)} \mid \underline{x}_{-1}{}^{(j)}) \cdot \mathrm{P}(\underline{x}_{-1}{}^{(j)} \mid \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2}) \qquad (3)$$

$$\text{with} \quad C = \frac{1}{\sum_{l=0}^{2^M-1} \mathrm{P}(\hat{\underline{x}}_0 \mid \underline{x}_0{}^{(l)}) \cdot \sum_{j=0}^{2^M-1} \mathrm{P}(\underline{x}_0{}^{(l)} \mid \underline{x}_{-1}{}^{(j)}) \cdot \mathrm{P}(\underline{x}_{-1}{}^{(j)} \mid \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2})} \quad .$$

see 4.2b)

$\mathrm{P}(\underline{x}^{(i)} \mid \hat{\underline{x}}, ...)$ of the 2^M different parameter values $v^{(i)}$ (or bit patterns $\underline{x}^{(i)}$), $i = 0, 1, ..., 2^M - 1$, using reliability information DRI as well as a priori knowledge about the distribution and (if required) the correlation of adjacent parameter values. In the second step the parameter $v^{(\text{est})}$ is estimated. This estimate is not restricted to the 2^M discrete entries $v^{(i)}$ of the quantization table used at the encoder.

In contrast to the various source controlled channel decoding techniques the main difference is that the optimization tries not to minimize the bit error rate, but to minimize some parameter error criterion. This kind of error criterion can have a closer match to the subjective performance. Parameter-individual estimators are possible. For most of the parameters the minimum mean square error criterion (MS) is adequate whereas the MAP criterion should be taken for some specific parameters such as the pitch lags.

4.2 Calculation of A Posteriori Probablities

a) A priori knowledge of the parameter distribution

For the estimation of speech codec parameters the *a posteriori probability* about any transmitted parameter index i is required. The following relation can be proven [9]

$$\mathrm{P}(\underline{x}^{(i)} \mid \hat{\underline{x}}) = C \cdot \mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(i)}) \cdot \mathrm{P}(\underline{x}^{(i)}) \qquad (4)$$

$$\text{with} \quad C = \frac{1}{\sum_{l=0}^{2^M-1} \mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(l)}) \cdot \mathrm{P}(\underline{x}^{(l)})} \quad .$$

Varying the a posteriori term over i, we get the conditional probability for each possibly transmitted combination $\underline{x}^{(i)}$ provided that $\hat{\underline{x}}$ had been received. The source dependent term $\mathrm{P}(\underline{x}^{(i)})$ is called the *0th order a priori knowledge*, because it consists of a simple histogram of $v^{(i)}$. This histogram can be measured in advance using representative speech signals.

If there is no knowledge available about the source statistics, one can only exploit the DRI in terms of $\mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(i)})$ assuming the parameters $v^{(i)}$ being equally likely. In this case eq. (4) simplifies to

$$\mathrm{P}(\underline{x}^{(i)} \mid \hat{\underline{x}}) = \frac{\mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(i)})}{\sum_{l=0}^{2^M-1} \mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(l)})} \quad . \qquad (5)$$

In practice, this simplification does not hold very well because we use e.g. Lloyd-Max quantizers yielding identical quantization error variance contributions of each quantization interval i but non-identical probabilities $\mathrm{P}(\underline{x}^{(i)})$. If the codec parameter can be expressed according to eq. (1), then eq. (5) and MS estimation are equivalent to applying eq. (2).

b) A priori knowledge of the parameter correlations

The classical approaches of speech encoding aim at minimizing the residual redundancy of codec parameters. However, due to practical reasons residual correlations between successive speech codec parameters can be observed mostly. As already mentioned by Shannon [19] this source coding sub-optimality should be exploited at the receiver. In contrast to [14], we exploit the residual redundancy not per bit but per parameter. As shown in [9] the a posteriori term can easily be extended to regard as well parameter correlations: The maximum information that is available at the decoder consists of the complete history of received bit combinations according to

$$\begin{aligned} &\mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{x}}_{-1}, \hat{\underline{x}}_{-2}, ..., \hat{\underline{x}}_{-n}, ..., \hat{\underline{x}}_{-N}) \\ &\qquad = \mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2}) \end{aligned} \qquad (6)$$

with $\hat{\underline{x}}_{-n}$ denoting the bit combination n time instants[1] before the present one and N being the number of already received bit combinations in the past. The term $\hat{\underline{X}}_{-2} = [\hat{\underline{x}}_{-2}, \hat{\underline{x}}_{-3}, \hat{\underline{x}}_{-4}, ...]$ includes the complete history up to the received bit combination two time instants before the present one.

To compute the a posteriori term (6) it is necessary to find a statistical model of the sequence of quantized parameters. It seems reasonable to discuss the sequence of quantized parameters in terms of a Markov process of 1st order, i.e.

$$\mathrm{P}(\underline{x}_0 \mid \underline{x}_{-1}, \underline{x}_{-2}, ..., \underline{x}_{-N}) = \mathrm{P}(\underline{x}_0 \mid \underline{x}_{-1}) \; .$$

Solutions for higher order models can be derived in a similar manner [8].

Equation (6) can be transformed into the recursion of eq. (3) [9]. To emphasize that correlations between adjacent parameters are regarded, $\mathrm{P}(\underline{x}_0{}^{(i)} \mid \underline{x}_{-1}{}^{(j)})$ is called *1st order* a priori knowledge. In eq. (3) the term $\mathrm{P}(\underline{x}_{-1}{}^{(j)} \mid \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2})$ is nothing else but the resulting a posteriori probability $\mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{X}}_{-1})$ from the

[1] The term "time instant" denotes any moment when the regarded parameter is received. In the ADPCM codec e.g. it equals a sample instant, in CELP coders it may be a frame or a subframe instant.

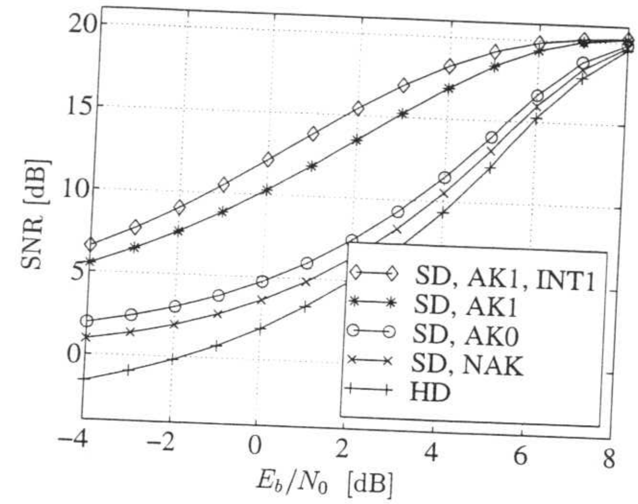

Figure 5: SNR performance of error concealment without channel coding (LMQ, $M = 4$, $\rho_{\tilde{v}\tilde{v}} = 0.9$, MS estimation)
See Table 1 for explanation of the legend

SD	Soft decision source decoding using ...
AK1	1st order a priori knowledge in eq. (3)
AK0	0th order a priori knowledge in eq. (4)
NAK	no a priori knowledge, eq. (5)
INT1	interpolation by exploiting one future bit combination, eq. (7)
FR	Frame repetition mechanism
HD	Hard decision source decoding by table lookup

Table 1: Abbreviations

previous time instant. Thus the a posteriori probabilities of all 2^M possibly transmitted bit combinations can be computed by a simple recursion, exploiting the maximum knowledge available at the decoder.

If at least a single-sample delay is allowed, interpolation instead of extrapolation can be performed by using

$$\mathrm{P}(\underline{x}_0{}^{(i)}|\hat{\underline{x}}_{+1},\hat{\underline{x}}_0,\hat{\underline{X}}_{-1}) = C \cdot \mathrm{P}(\underline{x}_0{}^{(i)}|\hat{\underline{x}}_0,\hat{\underline{X}}_{-1}) \cdot \sum_{l=0}^{2^M-1} \mathrm{P}(\hat{\underline{x}}_{+1}|\underline{x}_{+1}{}^{(l)}) \cdot \mathrm{P}(\underline{x}_{+1}{}^{(l)}|\underline{x}_0{}^{(i)}) . \quad (7)$$

The term $\mathrm{P}(\underline{x}_0{}^{(i)}|\hat{\underline{x}}_0,\hat{\underline{X}}_{-1})$ is taken from recursion (3). In the case of residual parameter correlations interpolation provides a further SNR gain relative to extrapolation.

4.3 Individual Parameter Estimation

For most speech codec parameters the minimum mean square error criterion (MS) is appropriate. These parameters may be PCM speech samples, spectral coefficients, gain factors, etc. In contrast to that some parameters as e.g. the pitch period require a different error criterion. The simplest one is the MAP (maximum a posteriori) estimator.

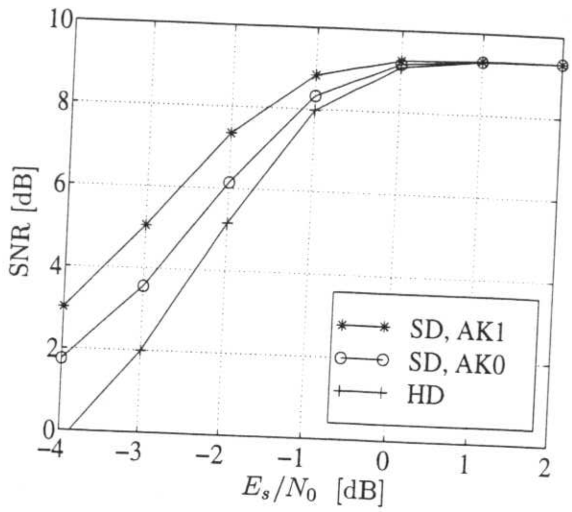

Figure 6: SNR performance of error concealment concatenated to error correction using the algorithm of Bahl et al. [6] for channel decoding (LMQ, $M = 2$, $\rho_{\tilde{v}\tilde{v}} = 0.9$, MS estimation), $E_s = 2E_b$ because of $r = 1/2$
See Table 1 for explanation of the legend

a) The MAP estimation

The MAP estimator requires the least additional computational complexity. It follows the criterion

$$v^{(\mathrm{MAP})} = v^{(\nu)} \quad \text{with}$$
$$\mathrm{P}(\underline{x}_0{}^{(\nu)} \mid \hat{\underline{x}}_0, ...) = \max_i \mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, ...),$$
$$i = 0, 1, ..., 2^M{-}1 . \quad (8)$$

$\mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, ...)$ denotes any of the a posteriori probabilities given by eqs. (3), (4), (5), or (7) dependent on the chosen model order and the available a priori knowledge. The optimum decoded parameter in a MAP sense $v^{(\mathrm{MAP})}$ always equals one of the codebook/ quantization table entries. Regardless the type of the speech codec parameter, a MAP estimation always minimizes the decoding error probability [18].

b) The MS estimation

The optimum decoded parameter $v^{(\mathrm{MS})}$ in a mean square sense equals

$$v^{(\mathrm{MS})} = \sum_{i=0}^{2^M-1} v^{(i)} \cdot \mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, ...) . \quad (9)$$

According to the orthogonality principle of linear mean square (MS) estimation (see e.g. [18]) the variance of the estimation error $e^{(\mathrm{MS})} = v^{(\mathrm{MS})} - v$ is simply

$$\sigma^2_{e(\mathrm{MS})} = \sigma^2_v - \sigma^2_{v(\mathrm{MS})} \quad (10)$$

with σ^2_v being the variance of the undisturbed parameter v and $\sigma^2_{v(\mathrm{MS})}$ denoting the variance of the estimated parameter $v^{(\mathrm{MS})}$. Because of $\sigma^2_{e(\mathrm{MS})} \geq 0$ the variance of the estimated parameter is smaller than or equals the variance of the error free parameter. For the worst case channel with $p_e = 0.5$ the a posteriori probability degrades to $\mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, ...) = \mathrm{P}(\underline{x}_0{}^{(i)})$. As a consequence, the MS estimated parameter according to eq. (9) becomes zero if the quantization table entries $v^{(i)}$ as well as $\mathrm{P}(\underline{x}_0{}^{(i)})$ are distributed symmetrically around zero. Thus the MS estimation of the gain factors results in an inherent muting mechanism providing a graceful degradation of speech. This is one of the main advantages of soft decision speech decoding.

4.4 Simulation Results

The significant improvements which can be achieved by error concealment are demonstrated by the examples of Fig. 5 and Fig. 6. In both cases the output of the source encoder is modelled by the quantized samples of a first order Gaussian Markov process with correlation $\rho_{\bar{v}\bar{v}} = 0.9$. Lloyd-Max quantizers (LMQ) with M bits per sample are used. The available total bit rate is assumed to be 4 bit per sample.

a) No channel coding

We choose $M = 4$ and use the total bit rate for source encoding. For clear channel conditions the SNR of the decoded parameter amounts to 20.22 dB. Fig. 5 shows that error concealment by soft decision source decoding (SD) with different degrees of a priori knowledge can achieve a significant quality improvement in comparison to conventional hard decision source decoding (HD) by table lookup.

b) Soft decision source decoding after channel decoding

In this case we choose $M = 2$ and use the remaining bit rate for convolutional channel encoding with rate $r = 1/2$ and constraint length $L = 5$. For channel decoding the algorithm of Bahl et al. [6] is applied providing a soft decision output DRI (see Fig. 3). Due to the lower bit rate for source encoding the SNR of the decoded parameter cannot be higher than 9.3 dB, even if the channel is free of errors. In comparison to hard decision source decoding (HD) soft decision source decoding (SD) can provide a distinct improvement, depending on the kind of available a priori knowledge.

The comparison of Fig. 5 and Fig. 6 shows that in this example soft decision source decoding with a priori knowledge of 1st order (SD,AK1) but without channel encoding could provide a better quality than soft decision source decoding with channel encoding. It should be kept in mind that usually the residual correlations are reduced by predictive coding, leading to an SNR higher than 9.3 dB under clear channel conditions. Nevertheless, *if* there are correlations remaining after source coding, and no attempt is (or can be) made to exploit them in the channel decoder, they might be better used during soft decision source decoding in a transmission system that assigns less of the bit rate to channel coding.

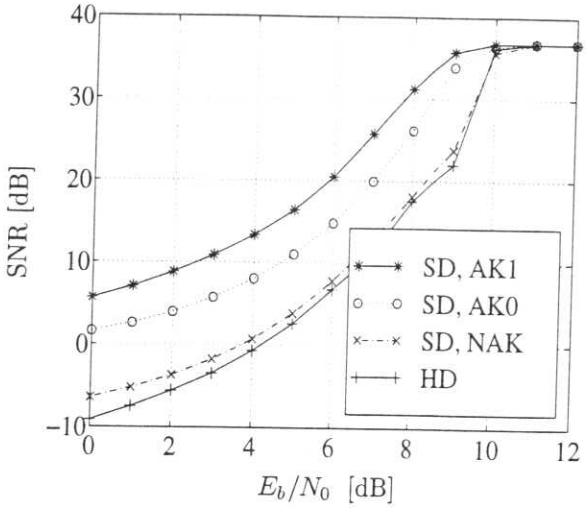

Figure 7: SNR performance of soft decision A-law PCM speech decoding: DRI to any bit, no channel coding
See Table 1 for explanation of the legend

5 Application Examples

5.1 PCM

We simulated a PCM transmission without channel coding over an AWGN channel with coherent BPSK demodulation. The DRI is assumed to be available to any received hard bit. Fig. 7 shows four different cases in terms of speech SNR as a function of the E_b/N_0 ratio.

In any case, the MS estimated speech degrades asymptotically to 0 dB with decreasing E_b/N_0, i.e. the inherent muting mechanism of MS estimation. The shape of the curves strongly depends on the order of exploited a priori knowledge. In comparison to conventional decoding (HD) the MS estimation without a priori knowledge (NAK) leads to a small speech SNR gain of about 1 ... 2 dB, while the exploitation of a priori knowledge allows gains of up to 10 dB (0th order), and up to 15 dB (1st order), respectively. This corresponds to a significant enhancement of speech quality although the model of the speech as Markov process of 1st order is still a very simple one and can be refined further [11].

5.2 ADPCM

To enhance the error concealment in DECT systems we designed a soft decision speech decoder for the G.726 ADPCM standard [1]. The 4 bit codec parameter is modelled as Markov process of 0th order according to eq. (4). A MS estimation of the residual signal sample instead of the quantization table entry is performed. The reconstruction of the recursively computed scale factor y is very sensitive to bit errors. Thus we performed an additional MS estimation of the term 2^y in advance of the residual signal estimation.

$[\frac{\text{bit}}{\text{parameter}}]$	LAR No.								LTP		RPE		
	1	2	3	4	5	6	7	8	Lag	Gain	Grid	Max.	Pulse
$H_0(\underline{x}_0)$	6	6	5	5	4	4	3	3	7	2	2	6	3
$H(\underline{x}_0)$	5.43	4.88	4.75	4.53	3.73	3.76	2.84	2.88	6.31	1.88	1.96	5.39	2.86
$H(\underline{x}_0 \vert \underline{x}_{-1})$	4.46	4.29	4.18	4.09	3.37	3.39	2.49	2.46	5.75	1.74	1.96	4.29	2.86

Table 2: Perfect information content $H_0(\underline{x}_0)$, entropy $H(\underline{x}_0)$, conditional entropy $H(\underline{x}_0 \mid \underline{x}_{-1})$

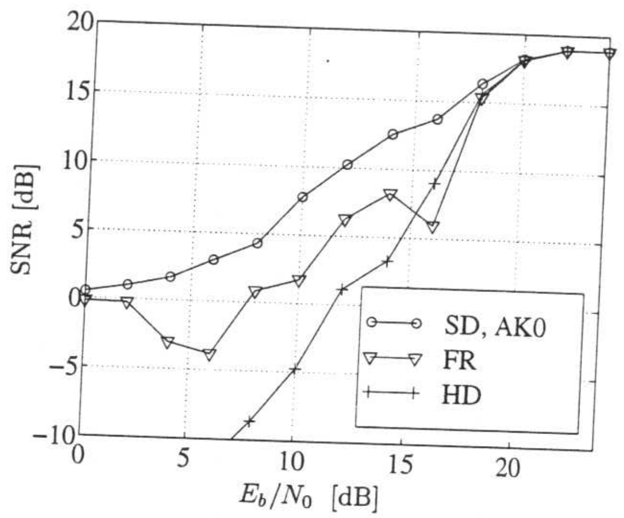

Figure 8: SNR performance of soft decision ADPCM speech decoding: one DRI to any DECT frame of 10 ms (320 bits), no channel coding See Table 1 for explanation of the legend

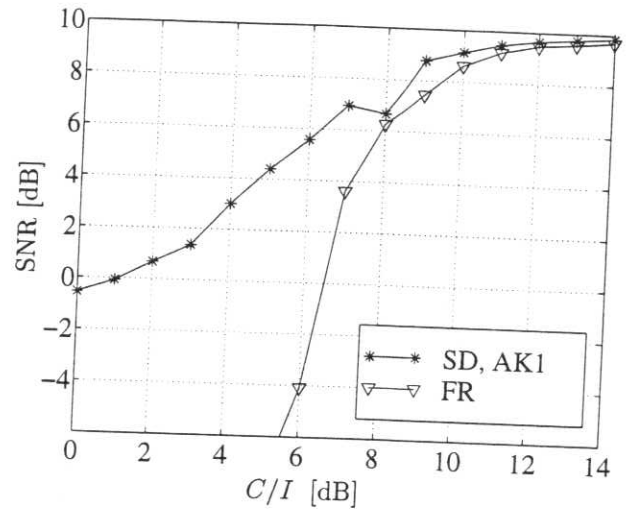

Figure 9: SNR performance of soft decision GSM full rate speech decoding: DRI to any bit, TU50 channel
See Table 1 for explanation of the legend

To simulate fading situations similar to channel characteristics of the DECT system we performed a DPSK modulation over a non-frequency selective fading channel with 2-path selection diversity. The frame synchronization was assumed to be perfect. The DRI is available only once per frame of 320 bits. Fig. 8 shows the simulation results indicating a significant superiority of soft decision speech decoding over hard decision speech decoding. As reference a simple frame repetition mechanism with muting (FR) was simulated. In comparison to that the speech quality degrades much smoother by soft decision speech decoding.

5.3 GSM Full Rate Codec

We applied the above described conception of soft decision speech decoding to the GSM full-rate speech decoder [2]. For simplicity, each codec parameter has been modelled as a Markov process of 1st order, although the entropy measurement of Table 2 shows that for the two parameters RPE grid and RPE pulse a model order of zero would be sufficient.

We found out that the non-integer GSM codec parameters can be well estimated using a MS estimator. In contrast to that, the estimation of a pitch period or the RPE grid position should be performed by a MAP (maximum a posteriori) estimator.

In Fig. 9 the results of a complete GSM simulation using the COSSAP GSM library [5] with speech and channel coding, interleaving, modulation, a channel model, demodulation and equalization are depicted. The channel model represents a typical case for an urban area (TU) with 6 characteristic propagation paths [4] and a user speed of 50 km/h (TU50). Soft decision output channel decoding is carried out by the algorithm of Bahl et al. [6]. The reference conventional GSM decoder performs error concealment by a frame repetition (FR) algorithm as proposed in [3]. The bad frame indicator is simply set by the evaluation of the CRC.

The SNR surely is not the optimum measure for speech quality. However, informal listening tests show the significant superiority of the soft decision speech decoder in comparison to the conventional decoding scheme in all situations of vehicle speeds and C/I ratios. The new error concealment technique provides quite a good subjective speech quality up to C/I = 6 dB, whereas the conventional frame repetition produces severe distortions already at C/I = 7 dB. Even long error bursts caused by a low vehicle speed can be decoded sufficiently by the new technique. In the soft decision decoding simulation, the hard annoying clicks in the case of CRC failures and the synthetic sounds of the frame repetition disappeared completely and turned into a slightly noisy or modulated speech. This gives a significant enhancement of the speech quality.

6 Conclusions

A first and obvious conclusion is that channel decoding as well as speech decoding can be improved by using soft input quality information (CSI or DRI) in combination with a priori knowledge on bit level or parameter level. This is in line with Shannon's information theory and follows Hagenauer's postulation of "using soft decisions in all stages of a digital receiver" [13].

A substantial difference between soft decision channel decoding and soft decision speech decoding is the optimization criterion i.e. on the one hand the minimization of the bit error rate and on the other hand the minimization of some error criterion in the real-valued parameter domain. The latter criterion seems to be related more closely to the subjective acoustical perception than the bit error criterion.

However, several fundamental open questions remain:

- Which is the more efficient approach? Should we avoid bit errors or should we conceal parameter errors?
- To which extent can we exploit quality and a priori information twice, i.e. for improving channel decoding and for improving speech decoding as well?
- How much channel coding do we really need if we design the speech encoder such that it leaves deliberately residual redundancy to support the error concealment process?

These questions path the way to further investigations. A first step in this direction is done in [7, 8] by taking fundamental aspects of rate distortion theory into consideration.

Independent of these theoretical investigations it can be concluded that the described error concealment technique can be used as a compatible extension to any transmission standard as it does in this case

- not require any modifications of the transmitter and
- preserves bit exactness of the speech decoder in case of error free transmission.

7 References

[1] "CCITT Recommendation G.726, 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)", *ITU*, August 1990.

[2] "Recommendation GSM 06.10, GSM Full Rate Speech Transcoding", *ETSI/TC SMG*, February 1992.

[3] "Recommendation GSM 06.11, Substitution and Muting of Lost Frames for Full Rate Speech Traffic Channels", *ETSI/TC SMG*, February 1992.

[4] "Recommendation GSM 05.05, Radio Transmission and Reception", *ETSI/TC SMG*, February 1992.

[5] Synopsys Inc., Mountain View, CA, USA, "COSSAP — Model Libraries Volume 3", 1995.

[6] L.R. Bahl and J. Cocke and F. Jelinek and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate", *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.

[7] T. Fingscheidt and P. Vary and J.A. Andonegui "Robust Speech Decoding: Can Error Concealment be Better Than Error Correction?", *Proc. of 2. ITG-Fachtagung Codierung für Quelle, Kanal und Übertragung*, Aachen, Germany, March 1998.

[8] T. Fingscheidt "Softbit-Sprachdecodierung in digitalen Mobilfunksystemen", PhD thesis, to be published by P. Vary in *Aachener Beiträge zu digitalen Nachrichtensystemen*, vol. 9, 1998.

[9] T. Fingscheidt and P. Vary "Error Concealment by Softbit Speech Decoding", *Proc. of ITG-Fachtagung Sprachkommunikation*, Frankfurt a.M., Germany, pp. 7–10, September 1996.

[10] T. Fingscheidt and O. Scheufen, "Error Concealment in the GSM System by Softbit Speech Decoding", *Proc. of 9th Aachen Symposium "Signaltheorie"*, Aachen, Germany, pp. 229–232, March 1997.

[11] T. Fingscheidt and P. Vary, "Robust Speech Decoding: A Universal Approach to Bit Error Concealment", *Proc. of ICASSP'97*, Munich, Germany, vol. 3, pp. 1667–1670, April 1997.

[12] J. Hagenauer, "Vom Analogwert zum Bit und zurück", *Frequenz*, vol. 51, no. 9, pp. 211–227, September 1997.

[13] J. Hagenauer, "Soft-In/Soft-Out: The Benefits of Using Soft Decisions in All Stages of Digital Receivers", *Proc. of 3rd Int. Workshop on DSP Techniques Appl. Space Commun.*, ESTEC Noordwijk, The Netherlands, September 1972.

[14] J. Hagenauer, "Source- Controlled Channel Decoding", *IEEE Transactions on Communications*, vol. 43, no. 9, pp. 2449–2457, September 1995.

[15] J. Hagenauer and P. Hoeher, "A Viterbi Algorithm with Soft-Decision Outputs and its Applications", *Proc. of GLOBECOM*, Dallas, TX, USA, pp. 1680–1686, November 1989.

[16] S. Heinen and A. Geiler and P. Vary "MAP Channel Decoding by Exploiting Multilevel Source A Priori Knowledge", *Proc. of 2. ITG-Fachtagung Codierung für Quelle, Kanal und Übertragung*, Aachen, Germany, March 1998.

[17] T. Hindelang and A. Ruscitto "Kanaldecodierung mit a priori Wissen bei nicht binären Quellensymbolen", *Proc. of 2. ITG-Fachtagung Codierung für Quelle, Kanal und Übertragung*, Aachen, Germany, March 1998.

[18] J.L. Melsa and D.L. Cohn, "Decision and Estimation Theory", *McGraw-Hill Kogakusha*, 1978.

[19] C.E. Shannon, "A Mathematical Theory of Communication", *Bell Systems Technical Journal*, vol. 27, pp. 379–423, July 1948.